

Contents lists available at ScienceDirect

## Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

# Empathic accuracy of young boys and girls in ongoing parent-child interactions: Performance and (mis)perception



C

### Wang Ivy Wong<sup>a,b,c,\*,1</sup>, Wai Bong Patrick Tsui<sup>a,1</sup>, Tik-Sze Carrey Siu<sup>d</sup>

<sup>a</sup> Department of Psychology, University of Hong Kong, Hong Kong

<sup>b</sup> Gender Studies Program, Faculty of Social Science, The Chinese University of Hong Kong, Hong Kong

<sup>c</sup> Department of Psychology, The Chinese University of Hong Kong, Hong Kong

<sup>d</sup> Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong

#### ARTICLE INFO

Article history: Received 4 November 2019 Revised 20 October 2020 Available online 7 December 2020

Keywords:

Interpersonal interaction Performance estimation Social cognition Gender Empathic accuracy Parent and child

#### ABSTRACT

Understanding others accurately is crucial in relationships and learning. Research shows that adults face challenges in empathic accuracy, that is, the ability to read the content of others' moment-to-moment mental states during interactions. Although young children possess some empathic understanding, the extent of their empathic accuracy is uncharted. Using a new SSP, 106 Chinese children aged 60 to 80 months (M = 70 months) were assessed on their ability to infer the mental states of adults in ongoing parent-child interactions. Replicating and extending extant findings on adults and adolescents, the children's inferences were found to be, at least computationally on a scale of .00 to 1.00, more often inaccurate than accurate regardless of the gender of the targets or participants (overall accuracy rate = .28). However, both the children and their primary caregivers overestimated the children's performance. In addition, although the primary caregivers expected girls to outperform boys, no gender difference in empathic accuracy was found when controlling for verbal fluency. Drawing on the findings of this first-ever application of the empathic accuracy paradigm in young children, the implications of empathic accuracy performance and misperceptions about such accuracy are discussed.

© 2020 Elsevier Inc. All rights reserved.

\* Corresponding author.

- E-mail address: iwwong@cuhk.edu.hk (W.I. Wong).
- <sup>1</sup> Co-first author.

https://doi.org/10.1016/j.jecp.2020.105042 0022-0965/© 2020 Elsevier Inc. All rights reserved.

#### Introduction

The ability to correctly judge others' thoughts and feelings is important for interpersonal relationships and social learning, but it is unclear to what extent children possess this ability. Although studies have documented various forms of social understanding in children, little is known about the accuracy of their inferences about the content of others' specific mental states (broadly including intentions, desires, beliefs, knowledge, and emotions) in realistic interactions. Such ability is called empathic accuracy and represents the cognitive element of interpersonal accuracy (Ickes, 2001; Ickes & Hodges, 2013; Marangoni, Garcia, Ickes, & Teng, 1995).

Gleason, Jensen-Campbell, and Jckes (2009) showed empathic accuracy to be distinct from other empathic abilities (e.g., empathic concern) in preadolescents and adults, and thus it deserves to be studied as a construct in its own right. Higher levels of empathic accuracy are positively correlated with interpersonal skills in social interactions (Ickes, 2001, 2016; Ickes & Hodges, 2013) and support for the emotional needs of one's spouse (Verhofstadt, Buysse, Ickes, Davis, & Devoldre, 2008), but they are negatively correlated with dishonesty and dehumanization (Lee, Hardin, Parmar, & Gino, 2019) and aggression (Cohen, Schulz, Liu, Halassa, & Waldinger, 2015). A study of 116 adolescents (aged 10-14 years) in a school setting found those with greater empathic accuracy to have stronger friendships and to experience less relational victimization and fewer internalizing problems (Gleason et al., 2009). We argue that empathic accuracy is also required for effective social learning, a mechanism that is likely involved in many aspects of development such as gender roles and moral codes. Because the socialization of norms, attitudes, and appropriate behaviors takes place through both overt and covert instruction and body language (Blakemore, Berenbaum, & Liben, 2009), its effectiveness depends on the socialization agent's mental state being accurately interpreted. For example, children might learn to avoid playing with gender-atypical toys if they observe their parents frowning when they do so owing to their accurate interpretation of their parents' disapproval.

In the study reported here, we assessed the empathic accuracy of preschoolers using the standard stimulus paradigm (SSP), which requires participants to infer the specific content of others' moment-to-moment mental states in real-life interactions (lckes, 2016). To the best of our knowledge, this study constitutes the first research attempt to measure empathic accuracy in young children. We also assessed the children's self-perceptions and their primary caregivers' perceptions of the children's performance to explore whether the two groups had insight into such accuracy.

#### Children's empathy-related abilities

Research has identified signs of empathy from an early age (McDonald & Messinger, 2011) when empathy is defined loosely as a broad understanding of others' thoughts and emotions and/or the ability to respond with appropriate emotions and behaviors (Cuff, Brown, Taylor, & Howat, 2016). Although prior studies offer valuable knowledge on social cognition in young children, they do not address the accuracy of their assessments of moment-to-moment mental states in realistic, contextualized interactions. For example, infants have been shown to represent false beliefs (Onishi & Baillargeon, 2005) and emotional states (Hepach & Westermann, 2013) in interpreting behaviors and to make social judgments (Hamlin, Wynn, Bloom, & Mahajan, 2011). Given infants' limited cognitive and physical abilities, nonverbal tasks requiring an inference of social understanding based on primitive responses such as whether the infants like or dislike a stimulus or find a scenario surprising are generally used in such research. However, it is possible that the infants are responding to behavioral cues rather than representing others' minds (Vaish & Woodward, 2010), making interpretation a challenge.

Theory of mind (ToM) refers to the ability to infer another's mental state and use it to explain and predict behavior and is sometimes equated with cognitive empathy (McDonald & Messinger, 2011). ToM has been widely studied in preschool children, with research showing that 3-year-olds generally fail to understand false beliefs, whereas most 6-year-olds do (Wellman, Cross, & Watson, 2001).

However, ToM studies are largely confined to structured laboratory tasks (e.g., presentation of puppet stories) and might not reflect everyday abilities (Plastow, 2012).

Several other measures of empathy are similarly limited in ecological validity, and none provides an objective index of empathic accuracy. For example, adult reports of empathy in children are often administered using questionnaires (Li & Wong, 2016) presenting statements describing children exhibiting behavior presumed to reflect empathy (e.g., crying when an animal dies). Although such reports offer a snapshot of a child's empathic tendencies, they are subject to adults' viewpoint and recall biases and cannot provide a standardized index of accuracy.

Most methods of assessing young children's emotional understanding focus on the decoding of facial and gestural emotions. In the picture and story technique (Boyatzis, Chazan, & Ting, 1993), for example, children are presented with a hypothetical story and asked to indicate their own emotions or those of the target by pointing to relevant pictures or drawings. The choice-from-array task simply asks children to choose the person or puppet in a picture who expresses a particular emotion or to name a captioned emotion (Denham, Bassett, Brown, Way, & Steed, 2015). Although these methods can capture discrete correct or incorrect answers, the materials used are often static, hypothetical, or out of context, and the subtle changes in expressions that occur in natural dynamic interactions are often sacrificed (Ickes, 1997).

A method with greater ecological validity is to observe overt expressions of empathy by, for example, counting prosocial gestures inside or outside a laboratory (McDonald & Messinger, 2011). However, observation studies are time-consuming, difficult to replicate in large samples, and unsuitable when the research target is the accuracy of empathy rather than a tendency to display empathy. To investigate the accuracy of empathy directed toward the specific content of moment-to-moment mental states in dynamic interactions, a different approach, such as the empathic accuracy paradigms, is needed.

#### How accurate is most people's empathy?

Two paradigms have been used in most prior empathic accuracy research, which has largely involved adults (Ickes & Hodges, 2013). In the unstructured dyadic interaction paradigm, each pair of participants engage in an unstructured conversation, have the conversation videotaped, and are subsequently asked to infer each other's mental states using the video (e.g., Ickes, Bissonnette et al., 1990; Ickes, Stinson et al., 1990; Ickes & Tooke, 1988; Stinson & Ickes, 1992; Verhofstadt et al., 2008). In the SSP, genuine interactions (or a monologue or an interview in some cases) are prerecorded to form standardized videos, and participants infer the mental states of the unscripted video targets (e.g., Gesn & Ickes, 1999; Gleason et al., 2009; Marangoni et al., 1995). In both paradigms, participants' inferences of the targets' mental states at specific points are compared with the actual mental states reported by the targets. Both paradigms exhibit high inter-rater reliability (usually  $\alpha > .90$ ) and validity across settings and ages (e.g., empathic accuracy is lower in autistic adolescents with normal intelligence and correlates with perspective taking in adults and with internalizing problems in 10- to 14-year-olds) (Gleason et al., 2009; Ickes & Hodges, 2013; Marangoni et al., 1995; Stinson and Ickes, 1992), and two studies have reported good internal consistency in preadolescents and adolescents (Demurie, De Corel, & Roeyers, 2011; Gleason et al., 2009). Both paradigms also capture subtle changes in naturalistic dynamic interactions and allow mind-reading ability to be scored directly by crosschecking targets' mental states against those inferred by participants, thereby conferring ecological validity. However, the unstructured dyadic interaction paradigm allows empathic accuracy toward an actual interaction partner to be assessed, whereas the SSP allows assessments of individual differences through standardized stimuli (Gleason et al., 2009; Ickes & Hodges, 2013).

Empathic accuracy performance is measured on an index scale ranging from .00 (zero accuracy) to 1.00 (perfect accuracy). Across various interaction scenarios, the average accuracy score for adults ranges from ~.11 or .30 to .35 for married couples (the higher scores were for studies not relying on couples discussing conflicts) to ~.20 for strangers and ~.30 for close friends and therapist–client pairs (Gesn & Ickes, 1999; Ickes, 2011, 2016; Stinson & Ickes, 1992; Verhofstadt et al., 2008). Demurie et al. (2011) reported that the empathic accuracy in 18 healthy adolescents watching peer dyads acquainting and gaming fell within this range of performance. The high degree of consistency in relative

performance across targets (i.e., participants who exhibit superior performance on one target are also likely to do so on other targets compared with other participants) has been taken to suggest stable dispositional individual differences, although mean accuracy may vary across targets (Ickes & Hodges, 2013; Marangoni et al., 1995).

Mean empathic accuracy is affected by characteristics of the perceiver and target and the relationship between them (Ickes & Hodges, 2013). For example, expressive targets elicit stronger perceiver empathic accuracy (Zaki, Bolger, & Ochsner, 2008), and targets with whom one enjoys a close relationship are easier to read than strangers (Stinson & Ickes, 1992). Studies have also reported a female advantage in empathy in both children and adults, although gender differences are small or nonsignificant in performance-based and physiological measures of empathy (Blakemore et al., 2009; Eisenberg & Lennon, 1983). A meta-analysis of empathic accuracy (Ickes, Gesn, & Graham, 2000) suggests that studies showing female superiority are an exception, with gender differences due to differences in motivation and stereotype activation rather than empathic ability itself (Ickes et al., 2000; Klein & Hodges, 2001).

#### Perceptions of empathic accuracy

An accurate estimation of one's own empathic accuracy is critical to social decision making. For example, people who have a low degree of empathic accuracy and are aware of it may suppress overconfidence bias, seek more information, and adjust their inferences in subsequent interactions, thereby improving their social relationships and avoiding embarrassment. Among those with a high degree of empathic accuracy, in contrast, a correct self-perception may encourage them to make better use of their empathic ability.

However, Marangoni et al. (1995) and Ickes, Stinson et al. (1990) have shown that adults are unable to reliably estimate their empathic accuracy when asked "How well do you think you inferred [the video target's] thoughts and feelings [from not at all to extremely well]?" or to self-report an empathic accuracy measure. In the former study, the participants' insight into their own empathic accuracy was so lacking that their subjective postexperiment estimates did not correlate with their actual performance even when feedback was given throughout the task (Marangoni et al., 1995). Although neither study reported whether the incorrect estimations reflected overestimation, underestimation, or random guessing, other studies indicate that adults have a strong tendency to overestimate their abilities in general (Dunning, 2011).

No study to date has applied the empathic accuracy approach to children, but research suggests that children also tend to overestimate their abilities in many domains, ranging from intelligence and memory to social standing, possibly as an adaptive behavior that enhances optimism and motivation (Bjorklund & Blasi, 2012). Studies also show that students tend to exaggerate perceived gender differences in gender-stereotyped subjects such as math and art (Chatard, Guimond, & Selimbegovic, 2007) and that young children internalize unsubstantiated gender stereotypes about intelligence (Bian, Leslie, & Cimpian, 2017).

Vervoort et al. (2007) investigated adolescents' estimation of their parents' empathic accuracy toward them. Parents' perceptions of children's empathic ability may also be important because parent-child communication gaps can affect children's attachment (Leibowitz, Ramos-Marcuse, & Arsenio, 2002) and future social competence (Black & Logan, 1995). Another potential consequence of such gaps is ineffective social learning given that several channels of socialization (e.g., direct instruction, observational learning) require messages from the socialization agent to be accurately conveyed. It seems that parents and children alike tend to overestimate the latter's abilities (Miller, Manhal, & Mee, 1991).

Parental perceptions of children's abilities are influenced by gender (Eccles, Jacobs, & Harold, 1990; Furnham, Reeves, & Budhani, 2002), with parents tending to perceive boys as outperforming girls in stereotypically male domains such as math and sports (Eccles et al., 1990) and girls as outperforming boys in stereotypically female domains such as language and music (Eccles et al., 1990; Furnham et al., 2002). Parents have also been documented to hold gender biases in estimates of their own children's abilities (e.g., Eccles et al., 1990; Mondschein, Adolph, & Tamis-LeMonda, 2000).

#### The current study

We used the SSP to investigate young children's empathic accuracy, specifically toward adults in common dynamic parent-child interactions. We tested four hypotheses. First, in line with our first research question (i.e., how empathically accurate are preschool children in parent-child interactions?), we hypothesized that the young children in our study would score toward the lower end of the SSP scale.

Second, because few gender differences have been documented in performance measures of empathy (Eisenberg & Lennon, 1983) or empathic accuracy in adults (Ickes et al., 2000; Klein & Hodges, 2001), we hypothesized that boys and girls would not differ in such accuracy. We also did not hypothesize an effect of target gender because prior findings are inconclusive and seem to be confounded by target expressiveness (Klein & Hodges, 2001; Laurent & Hodges, 2009). Therefore, we included target gender as a factor and controlled for expressivity.

Third, we hypothesized that the children's self-estimations would not be associated with their performance, that children would overestimate their performance, and that girls' self-estimations would be higher than those of boys. This hypothesis was grounded in prior research showing that adults' ratings of their own empathic accuracy is uncorrelated with their performance (Ickes, Stinson et al., 1990; Marangoni et al., 1995), that overestimations of one's own abilities are common in both adults (Dunning, 2011) and children (Bjorklund & Blasi, 2012), and that children tend to expect gender differences in abilities (Bian et al., 2017; Chatard et al., 2007).

Finally, we hypothesized that the parent participants would overestimate their children's empathic accuracy as parents have been documented to do with children's cognitive performance at school (Miller et al., 1991). Furthermore, we predicted that parents would expect girls to be more accurate than boys because many parents hold gender-stereotypical expectations of children's abilities (Eccles et al., 1990; Mondschein et al., 2000).

#### Method

#### **Participants**

Participants were 106 children (58 girls and 48 boys) aged 60 to 80 months (M = 70 months) in kindergarten grade 3 (K3) and their primary caregivers (n = 106, of which 10 were fathers and 8 were unreported). This sample size has  $\geq 80\%$  power to detect small within-group differences and correlations and medium between-group differences at  $\alpha = .05$  (two-tailed) (Faul, Erdfelder, Lang, & Buchner, 2007). Boys ( $M_{age} = 70.46$  months, SD = 4.28) and girls ( $M_{age} = 69.90$  months, SD = 4.12) did not differ in age, t(104) = 0.69, p = .494. Participants were ethnically Chinese, were reported to have normal or corrected-to-normal vision, were fluent in Cantonese, and had no known developmental disorders. Recruitment took place through local kindergartens. The five participating kindergartens were from five sociodemographically diverse districts of Hong Kong. All kindergartens were local, coeducational, and not-for-profit; thus, they were representative of the majority of kindergartens in Hong Kong. Socioeconomic status of the sample, as reflected by school district income, was diverse. The median household income of the districts ranked from 6 to 18 out of 18 districts (Hong Kong Census and Statistics Department, 2018). The median household income of Hong Kong (HKD 26500) falls in the middle of the range of median household incomes of the five districts, which ranged from HKD 21100 to HKD 29900. The participation rate of the children was ~44%.

#### Empathic accuracy stimulus videos

Six parent-child dyads (three father-daughter dyads and three mother-daughter dyads) were recruited to produce stimulus videos. All dyads spoke the local language fluently, were ethnically the same as the participants, and involved children aged similarly as the participants (i.e., 60–80 months). The video recording followed the format of the SSP (e.g., Gesn & Ickes, 1999; Gleason et al., 2009; Klein & Hodges 2001; Marangoni et al., 1995) except that our videos featured parent-child

dyads. All dyads received the same instructions, namely that (1) they were to naturally engage in a casual conversation, (2) they were given a Kinder chocolate egg and were told that it was up to the parent to decide whether to use it in the interaction, and (3) they were told that they were free to use any materials in the room. The interaction during the video recording was then spontaneous and not scripted. Targets were not restricted to a specific topic of conversation as in other studies so as to encourage naturalism and diversify the coverage of everyday parent–child interactions. Recording took place in a room at the dyad's home, and persons other than the dyad and the researcher were asked to leave the scene. The three mother–child dyads engaged in verbal discussions, two father–child dyads engaged in a mix of play activity and discussions, and one father–child dyad had a verbal–only discussion. The average unedited video lasted for ~10 to 15 min.

Immediately after recording, the parent targets reviewed the recording and wrote down their mental states and the specific moments (time stamp) where they appeared. Each parent recorded on average ~10 mental states. Prior studies have not reported concerns about targets hiding their mental states from the researcher. No method of measuring the specific content of mental states can fully avoid this possibility, and the empathic accuracy method is one of the most objective and feasible methods available. The intention to hide mental states from the researcher is likely low because targets were not asked to report to acquaintances. In addition, the interaction scenarios were not sensitive and targets were assured of confidentiality.

The recordings were then edited to remove unwanted "disturbances." These included when the child was distracted or when the dyad interacted with objects or the researcher outside the camera, prolonged periods without mental states or interaction, and occurrences that could not be understood without seeing the rest of the video (Demurie et al., 2011). Each of the final six stimulus videos consisted of five randomly chosen mental states and lasted for ~ 3 to 4 min (M = 226 s, SD = 21). Because target expressiveness might affect empathic accuracy (Zaki et al., 2008), 14 pilot raters saw all the videos and rated the parents on expressiveness. A paired-samples *t* test showed that the fathers (M = 9.64, SD = 2.21) and mothers (M = 9.88, SD = 1.61) did not differ, t(13) = -0.44, p = .664. To test whether the six videos were comparable in expressiveness, we conducted a repeated-measures analysis of variance (ANOVA). The differences were not significant, F(5, 65) = 1.83, p = .119. Consistently, none of the pairwise comparisons was significant (all ps > .05). See online supplementary material for a description of the video content.

#### Procedures

Children were tested individually in a quiet room at school. Although six stimulus videos were produced, each child viewed four of them in order to keep the length of the procedures in check. Children were first assessed on verbal fluency, which served as a proxy of general cognitive ability for control purposes. They then viewed two randomly chosen stimulus videos (one father video and one mother video in random order), followed by a break during which they played with stickers on their own to reduce fatigue, and then viewed another two randomly chosen stimulus videos (again, one father video and one mother video in random order). Children then were interviewed about their perceived performance and, lastly, peer ratings of peer relations, which served to provide preliminary evidence of criterion validity that would corroborate prior evidence of validity. These procedures took ~30 to 40 minutes. Primary caregivers were sent a questionnaire to complete at home at their convenience about their perception of empathic accuracy performance of their children and of boys and girls in general. Each participant received a Starbucks coupon worth HKD 25, and each class teacher received HKD 200 for assisting.

#### Verbal fluency task

Verbal fluency was included as a covariate because verbal ability and executive function may confound empathic accuracy. Children completed the Cantonese categorical verbal fluency test adapted from Chan and Poon (1999). Practice trials included words such as *taxi* given as examples of the category "transportation." Children then were asked to produce as many words as possible in 1 min for each of the gender-neutral categories "animals" and "fruits." The verbal fluency score is the sum of correct words generated.

#### Peer relations

Empathic abilities correlate with peer relations. For example, a meta-analysis found that ToM and peer popularity correlated in children aged 32 to 120 months (Slaughter, Imuta, Peterson, & Henry, 2015). In the current study, peer relations were assessed using items adapted from prior studies. Peer ratings, rather than nominations, were collected because they are more reliable and tend to correlate more strongly with sociometric measures (Crick et al., 2006). The rating scale was a 4-point scale (1 = *never or almost never true* to 4 = *always or almost always true*), aided by the visual presentation of an increasing number of ticks. Each child received ratings from all other participating children at the school (ranging from 10 to 38), and the ratings were averaged for each item. Children from one school (n = 10) were excluded for this measure because several classes participated and too few children could provide ratings for classmates they knew. Each child was presented with a list of photos and names of the other participating classmates and rated each one on the items "I like to play with him/her," "I avoid playing with him/her" (reverse-coded), and "S/he is my best friend" ( $\alpha = .88$ ). The first question was designed to reflect peer acceptance, the second question peer rejection, and the last question close friendships (Rys & Bear, 1997). The peer relations score was the sum of the three items.

#### Empathic accuracy task

Participants were instructed as follows: "I will now show you videos of real parent-child interactions. At specific moments, the videos will be paused, and I want you to tell me what the parent in each video is thinking or feeling." Before the testing, a sample video was shown. When the video paused, the researcher gave a standard prompt: "What is the parent in the video thinking/feeling?" If the child attempted to infer the mental state of the parent target, the test commenced. If the child did not understand the task, the video was replayed. After the warm-up, the empathic accuracy task was administered in two blocks as described earlier. At the exact moment where the parent target had recorded any mental state, the video was paused and the child was asked, following the standard prompt, to describe the mental state of the parent. The child's response was recorded verbatim. Researchers did not give encouragement during the test phase. If the child answered "don't know," the test moved on. With the warm-up phase, children responded very quickly without the need for further prompts.

Scoring largely followed the standard procedures (Gesn & Ickes, 1999; Gleason et al., 2009; Ickes, Bissonnette et al., 1990; Klein & Hodges, 2001; Marangoni et al., 1995). All empathic inferences were scored by three raters on how similar they were to the content of the target-reported mental states on a 3-point scale (0 = essentially different or if the child said "I don't know," 1 = similar but not the same, and 2 = essentially the same) (see Table 1 for the coding scheme). Inter-rater reliability was  $\alpha$  = .948. The raters' ratings were averaged before computing an empathic accuracy index.

The empathic accuracy index was computed by summing the scores for all 20 empathic inferences (five for each of four videos viewed), dividing this total by 20, and then dividing that number by 2 (the maximum score for any given inference). Thus, participants with the maximum raw score (i.e., 40) had an index of 1.00, meaning that they were fully correct on all inferences; those with a 0 raw score had an index of .00, meaning that none of their inferences was correct.

#### Children's perception

A numeric-point scale was preferred to a scale with descriptive options (e.g., *very good, good*) when assessing performance estimates because it circumvents the problem of different participants using different anchors of what the descriptors mean. After watching the videos, children answered the following question: "Out of 100 points, what score do you think you got for guessing the thoughts and feelings of the parents?" The 100-point scale was used because it is more intuitive to children than the 40-point scale of the raw performance scores, but if children indicated difficulty in understanding the

Table	1
-------	---

EA coding scheme.

Target's mental state	Participant's inference	EA score (2, 1, or 0)
Example: I am thrilled because he knows the answer.	He is happy because he knows the answer. He is happy. He is happy because he is going to school. He is angry because he knows the answer.	2 1 1 1
	në is sau.	0

Note. EA, empathic accuracy.

scale of 100, they were asked to rate from 0 to 10. Both scales are directly comparable to the empathic accuracy index, which ranged from .00 to 1.00.

#### Primary caregivers' perception

Primary caregivers' perception about their estimation of children's performance was assessed in a format similar to that in prior studies of performance estimation (Halpern, Straight, & Stephenson, 2011). To give primary caregivers an anchor to make the estimates, they were provided with a description of the task design and also the coding scheme. The question appeared as follows: "In this study, we show K3 children several videos of parent–child dyads interacting during some common daily activities (e.g., playing a game, talking about school). At various points during the videos, we pause to ask the K3 children what they think the parents in the videos are thinking or feeling. There are a total of 20 instances where the child is asked to make inferences. These 20 instances could score a total of 40 points maximum (2 points for completely correct, 1 point for partially correct, 0 points for incorrect answers). What do you think the average score will be for (1) K3 boys, (2) K3 girls, and (3) your child? Please make a guess even if you are not sure."

#### Results

#### Empathic accuracy index

There were a total of 2120 empathic inferences. All study variables had skewness (<2) and kurtosis (<3) that met requirements for normal distribution (Kline, 2011). The raw scores of empathic accuracy ranged from 1.33 to 22.67 (M = 11.12, SD = 4.88). The index ranged from .03 to .57 (M = .28, SD = .12). Overall index scores are summarized in Table 2. Of the six correlation coefficients among performance on the four videos viewed, four were significant and two were not (see Table 3).

Corroborating prior studies, peer relations correlated positively with the index in boys,  $r_s(42) = .33$ , p = .033, but not in girls,  $r_s(54) = -.10$ , p = .458. This finding adds to previously found correlates of empathic accuracy and provides preliminary evidence that the measure also possesses criterion validity, apart from high ecological and face validity, in children.

#### Empathic accuracy, perceiver gender, and target gender

To determine whether empathic accuracy was moderated by participants' gender or targets' gender, a mixed-design analysis of covariance (ANCOVA) with target gender as a within-participants factor and participant gender as a between-participants factor was performed. Covariates were verbal fluency and age. Results revealed no significant differences between boys and girls, F(1, 102) = 1.105, p = .296,  $\eta_p^2 = .01$ . However, a main effect of target gender, F(1, 102) = 5.790, p = .018,  $\eta_p^2 = .05$ , showed that the index was higher for the father videos (M = .30, SE = .01) than for the mother videos (M = .26, SE = .02). Child gender and target gender did not interact, F(1, 102) = 1.43, p = .235,  $\eta_p^2 = .01$ . The covariate, verbal fluency, had a significant effect, F(1, 102) = 7.71, p = .007, but age did not, F(1, 102), = 1.25, p = .255, likely because age variations were small.

Children's perception of their own empathic accuracy

Table	2	

Mean EA indices.

	Boys ( <i>n</i> = 48)		Girls ( <i>n</i> = 58)	
Measure	М	SD	М	SD
Overall EA index EA with mothers as targets EA with fathers as targets	.276 .244 .300	.13 .15 .15	.280 .269 .291	.12 .15 .14

Note. Possible range: .00 to 1.00. EA, empathic accuracy.

#### Table 3

Correlations among performance on the videos viewed by each child.

Video	1	2	3	4
1	_			
2	.27 (.005)	-		
3	.27 (.006)	.18 (.062)	-	
4	.24 (.012)	.18 (.073)	.28 (.004)	-

Note. Values are correlation coefficients and p values. Each child participant viewed a random set of four videos.

To allow for comparison, children's estimates were converted to proportions that ranged from .00 to 1.00. Children's estimated scores did not correlate with their actual performance, r(106) = .112, p = .252. A paired-samples *t* test followed up on the direction of the difference. Children's estimates (M = .85, SD = .25) were significantly higher than their actual scores (M = .28, SD = .12), *t* (105) = 34.29, p < .001, d = 6.69. Boys (M = .83, SD = .27) and girls (M = .87, SD = .23) did not differ in their estimates, t(104) = 0.66, p = .508.

Caregivers' perception of children's empathic accuracy

Primary caregivers' estimates of children's performance (on a 40-point scale) were also converted to proportions that ranged from .00 to 1.00. Their estimates did not correlate with children's actual performance, r(106) = -.02, p = .885. A paired-samples t test showed that primary caregivers' estimates (M = .70, SD = .17) were significantly higher than children's actual scores (M = .28, SD = .12), t(105) = 20.37, p < .001, d = 3.98.

To analyze gendered perceptions, a paired-samples *t* test was conducted to compare primary caregivers' estimates on the questions for K3 boys in general and K3 girls in general. Their estimates were significantly higher for girls (M = .77, SD = .14) than for boys (M = .67, SD = .13), t(105) = 9.86, p < .001, d = 1.92. We also compared estimates on the "my child" question by caregivers of girls (M = .71, SD = .17) and caregivers of boys (M = .68, SD = .16), and this gender difference was not significant, t(104) = 1.131, p = .261. Table 4 shows caregiver estimates and actual empathic accuracy indices.

#### Discussion

Although many adult studies and several adolescent studies have assessed empathic accuracy, such studies are missing from the child literature. To the best of our knowledge, the study reported here is the first to examine empathic accuracy and with a method (SSP) designed to test the ability to read the content of moment-to-moment mental states in young children. The SSP provides a naturalistic, contextually rich, and ecologically valid measure of such mind-reading ability. Our findings on children's performance on the SSP contribute to research on the diverse range of empathic abilities in young children (McDonald & Messinger, 2011) and pose interesting questions about how to characterize those abilities. We replicated and extended the findings of prior research on adolescents and adults by showing that empathic accuracy is challenging to young children. Studies have shown performance to be inaccurate more often than accurate in the paradigms assessing empathic accuracy when situated in interaction scenarios relevant to the respective developmental stage (e.g., couple interactions

	Caregiver estimate (SD)	Actual empathic accuracy index (SD)
Boys	.67 (.13)	.28 (.13)
Girls	.77 (.14)	.28 (.12)
My child	.70 (.17)	.28 (.12)

#### Table 4

Comparison of primary caregivers' estimates and children's actual empathic accuracy.

#### Note. Possible range: .00-1.00.

for adults, student-teacher interactions for adolescents, parent-child interactions for young children), an observation that was consistent across target and participant gender in our study. Furthermore, we found that both the participating children and caregivers held fallacious perceptions given that they overestimated the former's empathic accuracy. Interestingly, the caregivers, but not the children, also overestimated the female advantage in empathic accuracy performance.

#### Young children's empathic accuracy toward adults in interactions

This study provides initial evidence suggesting that the SSP can be applied to study young children's empathic accuracy with a high degree of reliability and feasibility. The participating children performed above zero (M = .28) and in fact performed at a level similar to that of adults in other scenarios (e.g., Gesn & Ickes, 1999; Ickes, 2016; Ickes, Stinson et al., 1990; Marangoni et al., 1995; Stinson & Ickes, 1992). Prior studies have reported the chance level of empathic accuracy tasks to be about .05 (i.e., 5% accurate if random inferred and actual mental states are compared; Ickes, Stinson et al., 1990; Stinson & Ickes, 1992).

Although our findings suggest that preschool children have some ability to infer adults' mental states, they score far below the computational maximum of 1.00. To put the participating children's performance into context, an empathic accuracy index of .50 could mean that they were completely accurate about half of the inferences and completely inaccurate about the other half or that they were partially accurate about all the inferences. The index of .28 thus signaled that the children were, computationally on a scale from .00 to 1.00, inaccurate more often than accurate in reading minds, at least in the context of an adult interacting with a child of a similar age as themselves. Although no reliable effects have been reported for target gender, our child participants performed better when viewing fathers than when viewing mothers despite differences in expressivity between the two adult genders being ruled out by the pilot raters. One possibility is that the depicted mothers engaged in more conversation, whereas two of the three fathers engaged in more play, with the play itself potentially providing additional cues helping the children to make inferences. Nevertheless, the finding of greater inaccuracy than accuracy holds true across the two target genders, with the children's performance reaching no higher than .30 even for the father targets.

Improving empathic accuracy would be beneficial given the importance of accurate mind reading for healthy interpersonal relationships and effective interaction (Cohen et al., 2015; Gleason et al., 2009; Ickes, 2001, 2016; Lee et al., 2019; Verhofstadt et al., 2008). Low levels of empathic accuracy may be posited to contribute to common interpersonal conflicts such as divorce and the breakup of friendships, and in children they may imply that the majority of socialization messages that adults are trying to convey are not perceived by children as intended. If that is the case, and some research suggests that it might be, then adults' socialization efforts may prove to be ineffective. For example, parents' gender role attitudes and/or behaviors are only loosely correlated with children's gender-typed preferences, behaviors, and abilities (Turner & Gervai, 1995; Wong & Yeung, 2019). In addition to examining the role of empathic accuracy in interpersonal relationships, future studies should also test the role it plays in the learning of social norms (e.g., gender, morality) in typically developing children and whether samples characterized by norm deviation (e.g., gender-nonconforming children) are also characterized by low levels of empathic accuracy.

Studies conducted in older samples have noted cross-target consistency in empathic accuracy and suggested that empathic accuracy may be a stable trait, with some individuals simply being better perceivers than others (Marangoni et al., 1995). It is unclear whether the similarity between the empathic accuracy scores recorded in our study and those of adult studies is related to empathic

accuracy being a stable trait. Several cross-sectional studies using variants of the empathic accuracy paradigms have found such accuracy to differ by age, increasing from adolescence to young adulthood and then declining up to middle adulthood, although the age effects were small (e.g., Kunzmann, Wieck, & Dietzel, 2018). If empathic accuracy is a stable trait from preschool onward, then research should be able to determine whether there is a critical period before preschool and whether such accuracy can be trained. However, when different targets are used, cross-target consistency tends to decline (Ickes & Hodges, 2013). In our sample, performance across the four videos viewed was only moderately correlated (ranging from nonsignificant to r = .28), suggesting that characteristics of the targets (other than expressivity) or contexts might have played a role in individual children's performance.

We also documented wide variations in performance levels among participants, with the highest score being .57 and the lowest being .03. The many adult studies in this area provide opportunities to explore the correlates of inter- and intra-individual differences in empathic accuracy (Ickes & Hodges, 2013), and our study offers foundational knowledge to launch such an investigation in children. The study was not designed to test the causes of differences in empathic accuracy, but the available research suggests that interventions should consider perceiver qualities such as verbal fluency (this study), vocabulary and fluid intelligence (Kunzmann et al., 2018), context characteristics such as relevance (Richter & Kunzmann, 2010), and target qualities such as target familiarity (Gesn & Ickes, 1999) and expressivity and transparency (Ickes & Hodges, 2013).

Although there may be ample room for children's empathic accuracy to improve, we note that several caveats are in order with respect to describing our child participants' scores as low (Ickes & Hodges, 2013). First, an average empathic accuracy score of .28 is low on a scale from .00 to 1.00. but this interpretation presumes that a score of 1.00 is possible. As an anonymous reviewer pointed out, more than 30 years of research has shown that it is not and that in any given study the average score has never been as high as .50, and even individual scores almost never exceed .70. The children's performance was in fact comparable to the average scores of adolescents and adults in other studies when asked to infer the thoughts and feelings of a stranger or even of same-sex close friends. Considering this data-informed judgmental context as well as Ickes' (2011) speculation that the effective range of empathic accuracy scores has an evolution-based "ceiling" of ~ .60 (instead of the computational maximum of 1.00), it might be inappropriate to diminish the participating children's achievement by describing the average score as "low." Second, it is possible that the targets intentionally obscured the messages conveyed for their own benefit, for that of the perceivers, or for the relationship between them. Third, empathic accuracy can be used to do good (e.g., identify a partner's needs) or to do harm (e.g., manipulate for an ill purpose). It has even been speculated that empathic accuracy ranges adaptively such that empathy is not so accurate as to weigh down others' interests over one's own but is sufficiently accurate to function effectively with others (Ickes, 2016).

#### Perceptions of performance level

Overconfidence is commonly observed in both adults and children (Bjorklund & Blasi, 2012; Dunning, 2011). Ickes, Stinson et al. (1990) and Marangoni et al. (1995) reported that adults' subjective evaluation of their empathic ability does not correlate with their actual performance. In the current study, we further tested whether the similar inaccuracy in children's self-estimations was the result of overestimation, underestimation, or random guessing. Although the children's estimates did not correlate with their actual performance, the children did not make random guesses. Instead, they grossly overestimated their performance, with the average child predicting an index score of .85, whereas the mean performance score was .28.

Some researchers have argued that overconfidence enhances self-efficacy and promotes optimism and motivation to attempt new things. For example, overconfidence in children in some instances predicts better performance (Bjorklund & Blasi, 2012). In a cross-sectional study of dating partners and longitudinal study of college roommates, Swann and Gill (1997) identified a frequent disconnect between confidence and impression accuracy. They concluded that although such a disconnect is sometimes problematic, its pervasiveness raises the possibility that confidence, even if misplaced, can contribute to relationship quality because overconfidence protects against the distress that would result from learning about a partner's transgressions.

However, the benefits of accurately estimating one's own empathic accuracy seem self-evident, for example, in motivating one to seek sufficient information to make an accurate inference, avoid embarrassment, and improve social interactions. It is difficult to see how overestimating one's ability to read others, and thus continuing to misread them, could be beneficial. For example, married couples have a tendency over time to rely on their previously formed understanding of each other (Ickes & Hodges, 2013). Although that understanding may prove to be sufficient early in the relationship, relying on it may prevent couples from improving their understanding of each other as the relationship evolves over time. Overconfidence is one potential reason for the relatively similar levels of empathic accuracy recorded in children in this study and in adults in prior research (see Gesn & Ickes, 1999; Ickes, 2016; Ickes, Stinson et al., 1990; Marangoni et al., 1995; Stinson & Ickes, 1992).

As noted, the participating primary caregivers also overestimated the children's ability to understand adults. Although they were provided with only a description of the task and scoring procedures, the extent of their overestimation (d = 3.98) suggests that it was not merely an artifact of incomplete knowledge of the task design. In addition, prior studies using a similar method to elicit estimates found that adults can be accurate in estimating the direction of group differences even when given only a short description of the tasks involved (Halpern et al., 2011). The gap between adult expectations and child ability prompts adults to exhibit sensitivity to possible miscommunication. Early theorists suggested that sensitivity to whether a message is successfully delivered may be even more important than the accuracy or inaccuracy of communication (Watzlawick & Beavin, 1967). Parents who accurately estimate their children's cognitive abilities have been shown to have more capable children (Miller et al., 1991). Perhaps participants were misled by the context of classroom tests where it is usually possible for children to get full scores. By generalizing from this judgmental context, caregivers and children themselves may expect to score in the .80s on the empathic accuracy test, especially when they did not know that no one ever scored so high on empathic accuracy. Our findings suggest that adults should refrain from over-reliance on children's ability to perceive and validate information and should instead engage in open discussion and solicit feedback to ensure effective communication. Such practice may be particularly important when counseling children during major life-affecting events such as parental divorce and migration.

#### Perceptions of gender differences

Parents often hold gender biases about children's abilities (Eccles et al., 1990; Furnham et al., 2002). In accordance with those biases, we found the participating caregivers to hold gendered expectations that were not supported by the children's actual performance data. Although there was no difference in performance on the empathic accuracy task between boys and girls, when asked to estimate how well boys and girls in general would perform on that task, the caregivers estimated that girls would outperform boys, with a large d (1.92). It is possible that our sample was not large enough to detect a small gender difference in performance, but it did have  $\geq 80\%$  power to detect a medium-sized difference (d = .55). Therefore, the null finding of a gender difference in performance suggests that any gender difference, if present, was smaller than medium. In addition, if the gender difference that the caregivers predicted (d = 1.92) had materialized, then our sample would have had 100% power to detect it (Faul et al., 2007). In sum, our findings support the conclusion that the caregivers overestimated the gender-based performance difference.

Although adults are fairly accurate in estimating the direction of gender differences at the group level (Koenig & Eagly, 2014) and sometimes underestimate them (Halpern et al., 2011), the caregivers in our study were not so concerning children's empathic accuracy. Consistent with the findings of prior studies (e.g., Eccles et al., 1990; Mondschein et al., 2000), they tended to overestimate rather than underestimate gender differences in their children. It should be noted that when the caregivers were asked to estimate children's performance on the "my child" item and that estimate was compared between caregivers of daughters and caregivers of sons, no gender difference was found. Parents rely more on gendered perceptions when thinking about other people's children than when thinking about their own (Fisher-Thompson, 1993), and children also rely more heavily on gender as a

predictive and explanatory factor for other children than for themselves (Eisenberg, Murray, & Hite, 1982), likely because familiarity reduces their reliance on stereotypes.

However, caregivers' expectation of a gender difference in empathic accuracy in children in general can have consequences. According to the expectancy effects model of parental socialization (Eccles et al., 1990), parental expectations of greater empathic ability in girls than in boys may be communicated to their children through daily messaging, thereby affecting children's self-belief and motivation to engage in empathy-related activities. Parents' stereotyped expectations can also be reflected in the activities in which they engage (Fivush & Wang, 2005; Lytton & Romney, 1991), thereby furthering the stereotype that male individuals are lacking in empathy. Interestingly, the boys and girls in this study expected to receive similar scores, which contradicts past findings that children tend to exaggerate perceived gender differences in gender-stereotyped subjects (Chatard et al., 2007) and to internalize gender stereotypes about intelligence (Bian et al., 2017). It is possible that the channeling of societal expectations that girls exhibit greater empathic accuracy than boys does not take effect until a later stage of development.

#### Limitations and strengths

Several limitations should be noted. First, although videos of unscripted interactions provide rich contextual information, cameras cannot capture all the dynamics that the human eye does. Furthermore, because the camera was placed in a way to keep both interactants in view in this research, the participants might not have been fully focused on the targets. Second, as in similar prior studies, only a limited set of scenarios and activities could be shown in the videos.

Third, the reduction in sampling out of concern for age appropriateness (five inferences per target) likely resulted in substantially more measurement error for each target than has been the case in adult studies (e.g., Marangoni et al., 1995), thereby constraining the size of the cross-target correlations. However, the number of videos and targets we showed was comparable to that shown in prior studies. In addition, the extant research has assessed empathic accuracy primarily within the context of a restricted type of activity and discussion content (e.g., therapist and patient talk about illness, teacher and student talk about schoolwork), whereas the parents and children in our videos engaged in a broader range of activities and conversations. Thus, the content was more diverse and representative of common parent–child activities (see supplementary material). Although thought-level factors may vary with activities and affect empathic accuracy (Ickes & Hodges, 2013), investigating thought-level variations requires a different study design. Such investigation would be a fruitful direction for future research (Ickes & Hodges, 2013).

Fourth, the empathic accuracy assessed in this study was directed at adults interacting with other children, not at adults interacting with the participants themselves, which is an inherent limitation arising from use of the SSP (Gleason et al., 2009; Ickes & Hodges, 2013). Children are likely to be more accurate in making inferences about their own parents or other adults with whom they interact. However, it is worthy of note that older samples do not exhibit greater empathic accuracy when inferring the mental states of interaction partners (Ickes, Stinson et al., 1990). Although our results cannot speak directly to children's empathic accuracy about an interaction partner, they are highly relevant to the development of empathic understanding because children often observe interactions between adult–child pairs (e.g., their teacher interacting with fellow classmates, their parents interacting with siblings).

Fifth, we did not attempt to establish correlations with other empathic skills because the literature suggests that a null finding might simply indicate that empathic accuracy is a distinct ability that should be studied independently (Gleason et al., 2009). However, as a measure of such accuracy about the content of specific mental states, the empathic accuracy paradigms, being both naturalistic and contextually meaningful, can be said to possess high degrees of face and ecological validity. There is also abundant evidence of the criterion validity of empathic accuracy in adults and young adolescents alike (Cohen et al., 2015; Gleason et al., 2009; Ickes, 2001, 2016; Lee et al., 2019; Verhofstadt et al., 2008). We did identify a correlation with peer relations in this study, at least in boys. As a first step toward examining empathic accuracy in children, our major aim was to describe both their level of performance and perception of performance. Thus, we leave further investigation of the construct's correlated outcomes during childhood for future studies.

Sixth, the children in all our stimulus videos were girls. Although the children in the videos were not the targets to be "read," and we did not observe any gender-specific discussions in the stimulus videos, the presence of girls alone might have rendered the context of the videos more favorable to girls than to boys. If that were true, however, then we should have found a significant female advantage, and we did not. In other words, the over-representation of girls in the videos provided a conservative test of the hypothesis that girls enjoy a performance advantage in empathic accuracy.

Seventh, although we controlled for age and verbal fluency, two important confounders of empathic ability, and also collected information on ethnicity, language fluency, school type, and school-level socioeconomic status to demonstrate the representativeness of our sample, we had only limited information on the backgrounds of the individual participants. The reliance on primary caregivers, rather than on parents alone, to estimate children's empathic accuracy could constitute an advantage because the primary caregiver is the person who knows the child best. Nevertheless, caution should be exercised when comparing our results with those of future studies relying on parent estimates, although nearly 90% of the caregiver sample comprised parents, making this a minor concern.

Finally, our findings might not generalize to other types of targets and age groups. The difficulty of being empathically accurate may depend on the interaction partners. For example, parents may be more direct and explicit with their children than with other adults. Alternatively, adult mental states may be guite complex for children to understand, and children may be better at judging other children because they are at a similar developmental level in terms of responding cognitively or emotionally to certain situations. Although the only relevant study reported no differences between using adults and using adolescents as targets (Gleason et al., 2009), more studies covering a wider age range of child perceivers and other types of targets (e.g., peer-peer) would be useful. In addition, to the extent that people from various cultures differ in gender differences and gender socialization (Wong & VanderLaan, 2020) and empathy (Cassels, Chan, Chung, & Birch, 2010), it would be interesting to conduct a similar study in a non-Chinese sample. Nevertheless, our use of a Chinese sample is valuable for two reasons. First, despite the potential cultural differences, our findings provide a useful reference for the study of empathic development. The general patterns of social abilities such as ToM (e.g., Wellman et al., 2001) and empathy (Chopik, O'Brien, & Konrath, 2016), and gender-related phenomena such as gender identity and gender stereotypes (e.g., Gibbons, 2000), are often culturally universal, with crosscultural comparisons revealing more similarities than differences. Second, given the predominance of samples from industrialized Western countries in developmental psychology (Nielsen, Haun, Kartner, & Legare, 2017), this research adds much-needed diversity and generalizability to the literature and is as representative as studies that rely on a single Western sample.

#### Conclusions

This article presents the first available evidence on the accuracy of young children's inferences about the mental states of adults during naturalistic dyadic interactions. Although their inferences are far from perfectly accurate, they were found to be as accurate as those of adults and adolescents in other research scenarios. Both the adults and children in this research were found to hold erroneous expectations. Not only did the children overestimate their level of accuracy, but so too did their primary caregivers, with the latter also overestimating the gender differences in children's performance. The empathic accuracy approach opens a new window on young children's moment-to-moment understanding of others' minds.

#### Acknowledgment

We thank Icy Yeung and Sylvia Shi for their help with data coding and the parents and children who participated or helped with the video production.

#### **Appendix A. Supplementary material**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2020. 105042.

#### References

- Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. Science, 355, 389–391.
- Bjorklund, D. F., & Blasi, C. H. (2012). Understanding self and others. In D. F. Bjorklund & C. H. Blasi (Eds.), Child and adolescent development: An integrated approach (pp. 268–303). Belmont, CA: Wadsworth.
- Black, B., & Logan, A. (1995). Links between communication patterns in mother-child, father-child, and child-peer interactions and children's social status. Child Development, 66, 255–271.
- Blakemore, J. E. O., Berenbaum, S. A., & Liben, L. S. (2009). The family as an agent of gender development. In J. E. O. Blakemore, S. A. Berenbaum, & L. S. Liben (Eds.), Gender development (pp. 271–304). New York: Psychology Press.
- Boyatzis, C. J., Chazan, E., & Ting, C. Z. (1993). Preschool children's decoding of facial emotions. Journal of Genetic Psychology, 154, 375–382.
- Cassels, T. G., Chan, S., Chung, W., & Birch, S. A. J. (2010). The role of culture in affective empathy: Cultural and bicultural differences. *Journal of Cognition and Culture*, 10, 309–326.
- Chan, A. S., & Poon, M. W. (1999). Performance of 7- to 95-year-old individuals in a Chinese version of the category fluency test. Journal of the International Neuropsychological Society, 5, 525–533.
- Chatard, A., Guimond, S., & Selimbegovic, L. (2007). "How good are you in math?" The effect of gender stereotypes on students' recollection of their school marks. Journal of Experimental Social Psychology, 43, 1017–1024.
- Chopik, W. J., O'Brien, E., & Konrath, S. H. (2016). Differences in empathic concern and perspective taking across 63 countries. Journal of Cross-Cultural Psychology, 48, 23–38.
- Cohen, S., Schulz, M. S., Liu, S. R., Halassa, M., & Waldinger, R. J. (2015). Empathic accuracy and aggression in couples: Individual and dyadic links. *Journal of Marriage and Family*, 77, 697–711.
- Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., & Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. Applied Developmental Psychology, 27, 254–268.
- Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. (2016). Empathy: A review of the concept. Emotion Review, 8, 144-153.
- Demurie, E., De Corel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders*, 5, 126–134.
- Denham, S. A., Bassett, H. H., Brown, C., Way, E., & Steed, J. (2015). "I know how you feel": Preschoolers' emotion knowledge contributes to early school success. Journal of Early Childhood Research, 13, 252–262.
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. Advances in Experimental Social Psychology, 44, 247–296.
- Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender difference. *Journal of Social Issues*, 46, 183–201.
- Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. Psychological Bulletin, 94, 100-131.

Eisenberg, N., Murray, E., & Hite, T. (1982). Children's reasoning regarding sex-typed toy choices. Child Development, 53, 81–86.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.

- Fisher-Thompson, D. (1993). Adult toy purchases for children: Factors affecting sex-typed toy selection. *Journal of Applied Developmental Psychology*, 14, 385–406.
- Fivush, R., & Wang, Q. (2005). Emotion talk in mother-child conversations of the shared past: The effects of culture, gender, and event valence. *Journal of Cognition and Development*, 5, 489–506.
- Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimation and estimations of their children's multiple intelligences. *Journal of Genetic Psychology*, 163, 24–39.
- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, 77, 746–761.
- Gibbons, J. (2000). Gender development in cross-cultural perspective. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 389–415). Mahwah, NJ: Lawrence Erlbaum.
- Gleason, K. A., Jensen-Campbell, L. A., & Ickes, W. (2009). The role of empathic accuracy in adolescents' peer relations and adjustment. Personality and Social Psychology Bulletin, 35, 997–1011.
- Halpern, D. F., Straight, C. A., & Stephenson, C. L. (2011). Beliefs about cognitive gender differences: Accurate for direction, underestimated for size. Sex Roles, 64, 336–347.
- Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. Proceedings of the National Academy of Sciences of the United States of America, 108, 19931–19936.
- Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. Journal of Experimental Child Psychology, 115, 16–29.
- Hong Kong Census and Statistics Department (2018). Population and household statistics analysed by district council district: 2017. Retrieved from https://www.statistics.gov.hk/pub/B11303012017AN17B0100.pdf.
- Ickes, W. (1997). Empathic accuracy. New York: Guilford.
- Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), Interpersonal sensitivity: Theory and measurement. LEA Series in Personality and Clinical Psychology (pp. 219–241). Mahwah, NJ: Lawrence Erlbaum.
- Ickes, W. (2011). Everyday mind reading is driven by motives and goals. *Psychological Inquiry*, 22, 200–206.
- Ickes, W. (2016). Empathic accuracy. In J. Hall, M. Schmid Mast, & T. West (Eds.), The social psychology of perceiving others accurately (pp. 52–70). Cambridge, UK: Cambridge University Press.
- Ickes, W., Bissonnette, V., Garcia, S., & Stinson, L. (1990). Implementing and using the dyadic interaction paradigm. In C. Hendrick & M. Clark (Eds.), Review of personality and social psychology. Vol. 11: Research methods in personality and social psychology (pp. 16–44). Newbury Park, CA: Sage.
- Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, 7, 95–109.
- Ickes, W., & Hodges, S. D. (2013). Empathic accuracy in close relationships. In J. A. Simpson & L. Campbell (Eds.), The Oxford handbook of close relationships (pp. 348–373). New York: Oxford University Press.

- Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. Journal of Personality and Social Psychology, 59, 730–742.
- Ickes, W., & Tooke, W. (1988). The observational method: Studying the interaction of minds and bodies. In S. Duck, D. F. Hay, S. E. Hobfoll, W. Ickes, & B. M. Montgomery (Eds.), Handbook of personal relationships: Theory, research, and interventions (pp. 79–97). Oxford, UK: John Wiley.
- Klein, K. J., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. Personality and Social Psychology Bulletin, 27, 720–730.
- Kline, R. B. (2011). Principles and practice of structural equation modeling. New York: Guilford.
- Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. Journal of Personality and Social Psychology, 107, 371–392.
- Kunzmann, U., Wieck, C., & Dietzel, C. (2018). Empathic accuracy: Age differences from adolescence into middle childhood. Cognition and Emotion, 32, 1611–1624.
- Laurent, S. M., & Hodges, S. D. (2009). Gender roles and empathic accuracy: The role of communion in reading minds. Sex Roles, 60, 387–398.
- Lee, J. J., Hardin, A. E., Parmar, B., & Gino, F. (2019). The interpersonal costs of dishonesty: How dishonest behavior reduces individuals' ability to read others' emotions. *Journal of Experimental Psychology: General*, 148, 1557–1574.
- Leibowitz, J., Ramos-Marcuse, F., & Arsenio, W. F. (2002). Parent-child emotion communication, attachment, and affective narratives. Attachment and Human Development, 4, 55–67.
- Li, Y.-H., & Wong, W. I. (2016). Gender-typed toy play and social abilities in boys and girls: Are they related? Sex Roles, 74, 399-410.
- Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, 109, 267–296.
- Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. Journal of Personality and Social Psychology, 68, 854–869.
- McDonald, N. M., & Messinger, D. S. (2011). The development of empathy: How, when, and why. In A. Acerbi, J. A. Lombo, & J. J. Sanguineti (Eds.), Free will, emotions, and moral actions: Philosophy and neuroscience in dialogue (pp. 333–361). Rome: IF Press.
- Miller, S. A., Manhal, M., & Mee, L. L. (1991). Parental beliefs, parental accuracy, and children's cognitive performance: A search for causal relations. *Developmental Psychology*, 27, 267–276.
- Mondschein, E. R., Adolph, K. E., & Tamis-LeMonda, C. S. (2000). Gender bias in mothers' expectations about infant crawling. Journal of Experimental Child Psychology, 77, 304–316.
- Nielsen, M., Haun, D., Kartner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. Journal of Experimental Child Psychology, 162, 31–38.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. Science, 308, 255-258.
- Plastow, M. (2012). "Theory of mind" II: Difficulties and critiques. Australasian Psychiatry, 20, 291–294.
- Richter, D., & Kunzmann, U. (2010). Age differences in three facets of empathy: Performance-based evidence. *Psychology and Aging*, 26, 60–70.
- Rys, G. S., & Bear, G. G. (1997). Relational aggression and peer relations, gender and developmental issues. Merrill-Palmer Quarterly, 43, 87–106.
- Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Development*, 86, 1159–1174.
- Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. Journal of Personality and Social Psychology, 62, 787–797.
- Swann, W. B., Jr., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners?. Journal of Personality and Social Psychology, 73, 747–757.
- Turner, P. J., & Gervai, J. (1995). A multidimensional study of gender typing in preschool children and their parents: Personality, attitudes, preferences, behavior, and cultural differences. Developmental Psychology, 31, 759–772.
- Vaish, A., & Woodward, A. (2010). Infants use attention but not emotions to predict others' actions. Infant Behavior and Development, 33, 79–87.
- Verhofstadt, L. L., Buysse, A., Ickes, W., Davis, M., & Devoldre, I. (2008). Support provision in marriage: The role of emotional similarity and empathic accuracy. *Emotion*, 8, 792–802.
- Vervoort, T., Crombez, G., Buysse, A., Goubert, L., DeBacker, T. D., & Ickes, W. (2007). Brief report: The accuracy of parents for the thoughts and feelings of their adolescent suffering from chronic fatigue: A preliminary study of empathy. *Journal of Pediatric Psychology*, 32, 494–499.
- Watzlawick, P., & Beavin, J. (1967). Some formal aspects of communication. American Behavioral Scientist, 10, 4-8.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. Child Development, 72, 655–684.
- Wong, W. I., VanderLaan, D. P. (2020). Early sex differences and similarities: Evidence across cultures? In F. M. Cheung & D. F. Halpern (Eds.), Cambridge Handbook of the International Psychology of Women (pp. 83–95). Cambridge University Press.
- Wong, W. I., & Yeung, S. P. (2019). Early gender differences in spatial and social skills and their relations to play and parental socialization in Hong Kong Chinese children. Archives of Sexual Behavior, 48, 1589–1602.
- Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, 19, 399–404.